Statistical Inference Project - Analysis of Tooth Growth for different supplements

Conrad Wong

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
library(datasets)
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(ggplot2)

# par(mfrow=c(1, 2))

table(ToothGrowth$supp, ToothGrowth$dose)

##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10

summary(ToothGrowth)

##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

There are 10 observations for each combination of dose (0.5mg, 1mg and 2mg) and supplement (Orange Juice and Vitamin C)

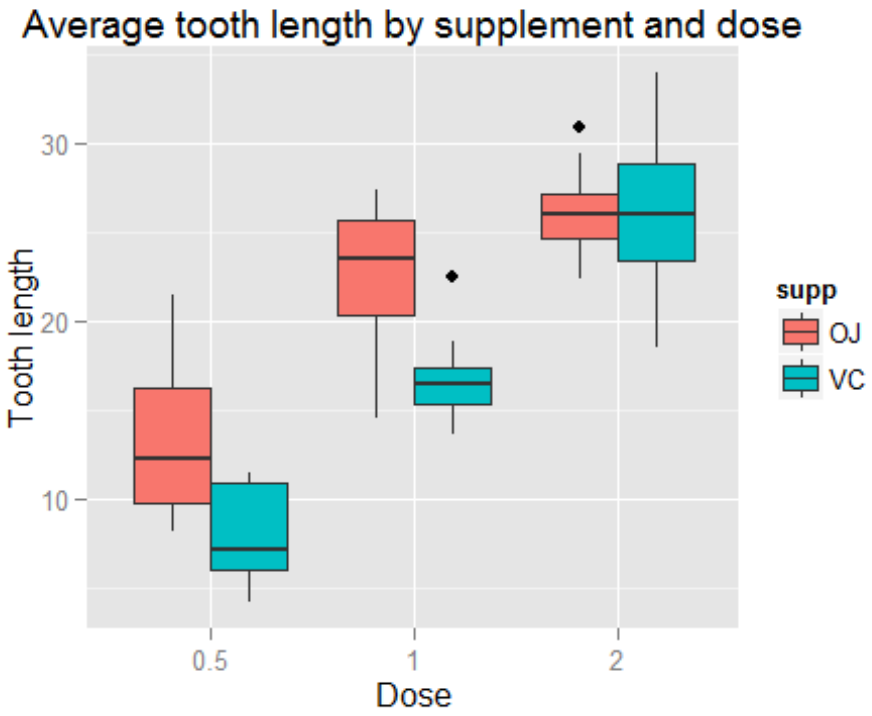
2. Provide a basic summary of the data.

```
ToothGrowth %>% group_by(dose, supp) %>% summarize(mean(len))

## Source: local data frame [6 x 3]
## Groups: dose
##
##   dose supp mean(len)
## 1  0.5   OJ     13.23
## 2  0.5   VC      7.98
## 3  1.0   OJ     22.70
## 4  1.0   VC     16.77
## 5  2.0   OJ     26.06
## 6  2.0   VC     26.14

g <- ggplot(ToothGrowth, aes(as.factor(dose), len)) + geom_boxplot(aes(fill=supp))
g <- g + labs(title="Average tooth length by supplement and dose", x="Dose", y="Tooth length")

g
```



From observing the summary table and the boxplot, it looks like tooth length increases as the dose increases. For doses 0.5mg and 1mg, it seems that Orange Juice (OJ) results in increased tooth length than Vitamin C (VC). For dose 2.0 there doesn't seem to be a big difference between both. We'll use hypothesis testing to validate these findings.

3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supplement and dose. (Only use the techniques from class, even if there's other approaches worth considering)

We'll test the following two hypothesis:

1) Dose of 2mg results in longer tooth than 1mg (hypothesis is "greater than")
2) Dose of 1mg results in longer tooth than 0.5mg (hypothesis is "greater than")

```
dose05 <- ToothGrowth[ToothGrowth$dose == 0.5, ]
dose1 <- ToothGrowth[ToothGrowth$dose == 1, ]
t.test(dose1$len, dose05$len, var.equal=FALSE, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  dose1$len and dose05$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605

dose2 <- ToothGrowth[ToothGrowth$dose == 2, ]

t.test(dose2$len, dose1$len, var.equal=FALSE, alternative = "greater")

##
##  Welch Two Sample t-test
##
## data:  dose2$len and dose1$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Conclusions:

1. Dose 1mg and 0.5mg comparison: The p-value of the test is 6.342e-08, so the null hypothesis is rejected (since it's less than alpha=5%), and we can conclude that the teeth length is greater for dose 1mg than dose 0.5mg. This can also be confirmed since 0 is not included in the confidence interval for the difference: 6.753323 to Inf

2. Dose 2mg and 1mg comparison: The p-value of the test is 9.532e-06, so the null hypothesis is rejected (since it's less than alpha=5%), and we can conclude that the teeth length is greater for 2mg than 1mg. This can also be confirmed since 0 is not included in the confidence interval for the difference: 4.17387 to Inf

Now, we'll test if there's a difference between the supplements within each dose:

1) Orange Juice vs. Vitamin C for dose 0.5
2) Orange Juice vs. Vitamin C for dose 1
3) Orange Juice vs. Vitamin C for dose 2

```
t.test(len~supp, var.equal=FALSE, data=dose05)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98

t.test(len~supp, var.equal=FALSE, data=dose1)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77

t.test(len~supp, var.equal=FALSE, data=dose2)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

Conclusions:

1. Dose 0.5mg: The p-value of the test is 0.006359, so the null hypothesis is rejected (since it's less than alpha=5%), and we can conclude that there is a difference between Orange Juice and Vitamin C for this dose.

2. Dose 1mg: The p-value of the test is 0.001038, so the null hypothesis is rejected (since it's less than alpha=5%), and we can conclude that there is a difference between Orange Juice and Vitamin C for this dose.

3. Dose 2mg: The p-value of the test is 0.9639, so we fail to reject the null hypothesis (since it's higher than alpha=5%), and we cannot conclude that there is a difference between Orange Juice and Vitamin C for this dose

4.State your conclusions and the assumptions needed for your conclusions.

Final Conclusion: As the dose increases, the theeth grow bigger. At dose 2mg, using Orange Juice or Vitamin C makes no difference. If you use dose 0.5mg or 1mg, using Orange Juice will result in bigger teeth.

Assumptions: The analysis assumes unequal variances, to be on the safe side. It also assumes data is approximately normal.